RAND

Assessing Research

The Researchers' View
Volume II

Steven Wooding, Jonathan Grant

RAND Europe

RAND

Assessing Research

The Researchers' View
Volume II

Steven Wooding, Jonathan Grant

Prepared for the
Higher Education Funding Council for England

RAND Europe

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

RAND® is a registered trademark.

Assessing Research: The Researchers' View: Volume II

Cover Image: A voting declaration on characteristics for research assessment systems.

Figure 1: Units of Assessment Represented

Figure 1: Units of Assessment Represented
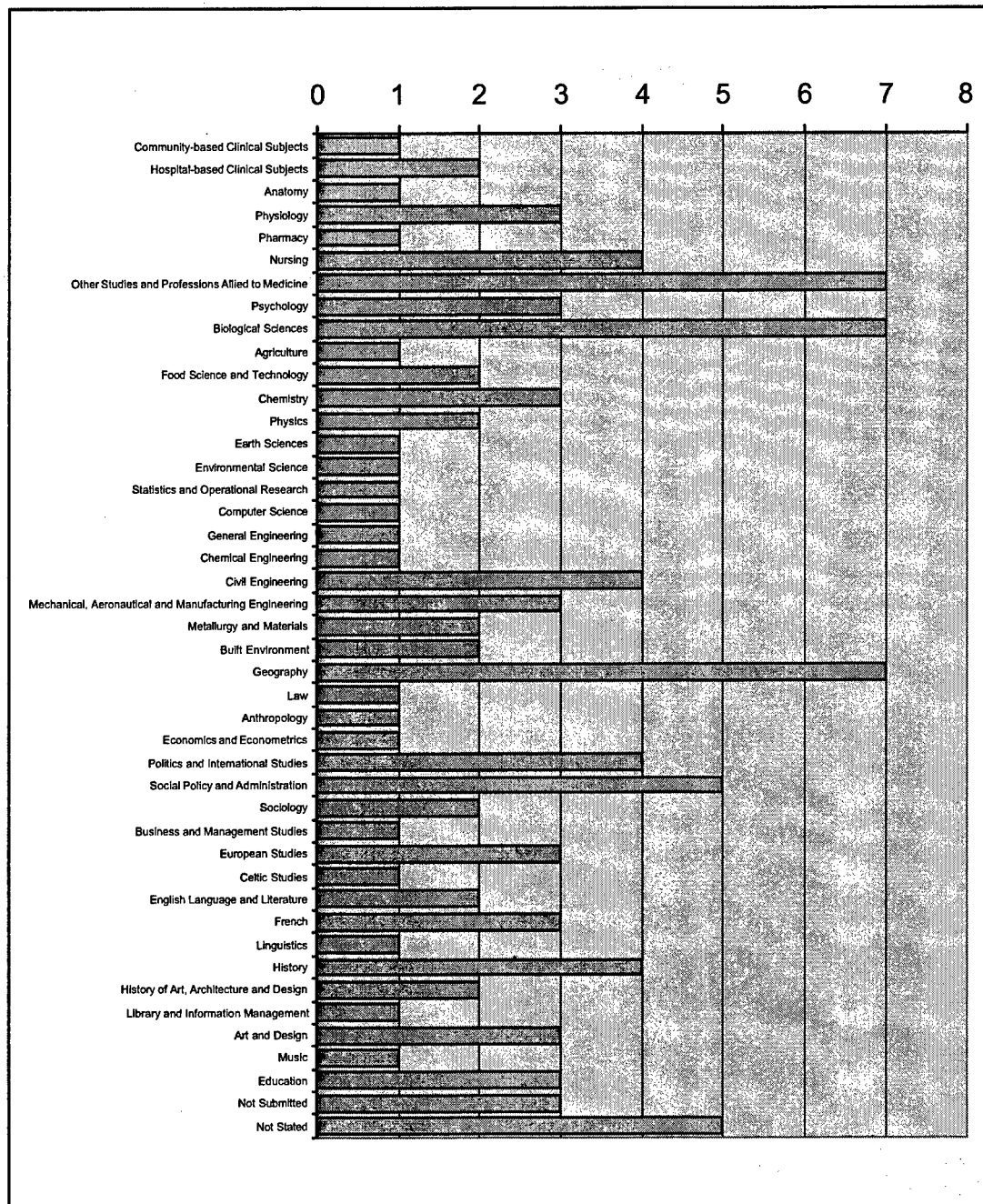
Annex A – Information on Participants

Table 1: Institutions Represented

Aberdeen University
Anglia Polytechnic University
Bath University
Bournemouth University
Bristol University
Brunel University
Cambridge University
Cardiff University
Chester College
Cranfield University
Durham University
Edge Hill College of Higher Education
Edinburgh College of Art
Exeter University
Glasgow Caledonian University
Glasgow University
Heriot Watt University
Imperial College
Kings College London
Lancaster University
Liverpool John Moores University
Liverpool University
London Metropolitan University
Luton University
Manchester University
Napier University
Newcastle University
Northumbria University
Oxford University
Plymouth University
Portsmouth University
Queen Mary's School of Medicine and Dentistry
Queen's University Belfast
Reading University

Royal Holloway University of London
Salford University
Southampton University
Southbank University
St Martin's College
Sussex University
Thames Valley University
The Bolton Institute
UMIST
University of Central Lancashire
University College London
University of East Anglia
University of Glamorgan
University of Kent at Canterbury
University of St Andrews
University of Sunderland
University of Surrey
University of Teeside
University of the West of England
University of Ulster
University of Wales, Aberystwyth
University of Wales Institute Cardiff
University of Wales Newport
University of Wales Swansea
University of Wales Bangor

Table 2: Unrepresented Units of Assessment

UoA No.	
1	Clinical Laboratory Sciences
4	Clinical Dentistry
5	Pre-clinical Studies
8	Pharmacology
17	Veterinary Science
22	Pure Mathematics
23	Applied Mathematics
29	Electrical and Electronic Engineering
31	Mineral and Mining Engineering
34	Town and Country Planning
41	Social Work
44	Accounting and Finance
45	American Studies
46	Middle Eastern and African Studies
47	Asian Studies
52	German, Dutch and Scandinavian Languages
53	Italian
54	Russian, Slavonic and East European Languages
55	Iberian and Latin American Languages
57	Classics, Ancient History, Byzantine and Modern Greek
58	Archaeology
62	Philosophy
63	Theology, Divinity and Religious Studies
65	Communication, Cultural and Media Studies
66	Drama, Dance and Performing Arts
69	Sports-related Subjects

Table 3: Participants by Role and Gender

		Role				
		Admin	Lecturer	Research	Senior	Total
Gender	female	16	8	1	13	38
	male	20	11	8	65	104
Total		36	19	9	78	142

Table 4: Research and Teaching Staff by Field and Role

		Field			Total
		Arts	Medicine	Social	
Role	Lecturer[1]	3	10	6	19
	Research[2]	1	8	0	9
	Senior[3]	18	40	20	78
Total		22	58	26	106

[1] Lecturer or Senior Lecturers with teaching and research responsibilities (principal lecturer, senior lecturer, lecturer)

[2] Research fellows with near exclusive research responsibilities

[3] Senior academics with large management responsibility (professor, head of department, reader)

Annex B – Task 1 Results

Two participants had to be excluded from the analysis because of a clerical error, and one participant did not vote. It should also be noted that a few participants did not cast all five of their votes.

Table 1: Characteristics of High Quality Research

Cluster	Characteristic	Workshop	Votes
Academic Impact	Academic Impact	4	11
Academic Impact	Academic Impact	5	10
Academic Impact	Academically influential	6	1
Academic Impact	Potential to have an impact	8	6
Collaborative	Collaboration	3	1
Collaborative	Themed and collaborative	6	2
Creative	Creativity	5	4
Creative	Creative	6	7
Creative	Creativity	8	8
Defining Research Agenda	Potential to move discipline forward	1	15
Defining Research Agenda	Highlights new questions	1	8
Defining Research Agenda	Moving the field forward	2	15
Defining Research Agenda	Agenda setting	4	9
Defining Research Agenda	Changing thinking	4	2
Defining Research Agenda	Defining an area	5	3
Defining Research Agenda	Defines research agenda	6	13
Defining Research Agenda	Creates debate	6	3
Defining Research Agenda	Advancement of field	7	5
Defining Research Agenda	Taking discipline forward	7	4
Defining Research Agenda	Agenda challenging	8	4
Defining Research Agenda	Stimulates further development	8	3
Defining Research Agenda	Developing new agenda	8	2
Defining Research Agenda	Agenda setting	9	13
High Quality Publication	High impact publication	3	6
High Quality Publication	Publication output	4	9
High Quality Publication	High quality publication	5	9
High Quality Publication	Output in terms of publication	7	8
Innovation	Innovative	1	8
Innovation	Innovation	3	6
Innovation	Innovation	4	7
Innovation	Innovation and boundary pushing	8	2
Innovation	Innovative	9	10
International Recognition	National and international impact	1	8
International Recognition	International Impact	3	6
International Recognition	International recognition	4	15
International Recognition	Internationally competitive	5	6
International Recognition	International recognition	6	11
International Recognition	Internationally leading	6	2
International Recognition	International impact	7	5
International Recognition	International peer recognition	8	9
International Recognition	National and internationally leading	9	7
Long Lasting	Long lasting	2	6
Long Lasting	Lasting impact	3	4
Long Lasting	Longevity	4	5
Long Lasting	Long lasting	6	3
Long Lasting	Longevity	7	3
Long Lasting	Longevity	9	4

Cluster	Characteristic	Workshop	Votes
Original	Originality	1	7
Original	Originality [4]	2	0
Original	Original findings	3	7
Original	Originality of method	3	2
Original	Originality and innovation	5	15
Original	Originality and novelty	6	14
Original	Original	7	8
Original	Novel application	8	2
Original	Original	9	6
Peer Recognition	Peer recognition	1	4
Peer Recognition	Peer/expert recognition	2	14
Peer Recognition	Peer recognition	4	14
Peer Recognition	Peer recognition	5	15
Peer Recognition	Peer recogniti on	9	4
Relevance	Relevance	4	2
Relevance	Fitness for purpose	5	3
Relevance	Relevant to and integrated with teaching	6	1
Relevance	Public relevance	8	2
Relevance	Empirical relevance	8	1
Research Culture	Research ethos	4	1
Research Culture	Research culture	5	7
Rigour	Rigorous	2	15
Rigour	Depth	2	9
Rigour	Rigour	3	5
Rigour	Methodological rigour	5	6
Rigour	Rigorous	6	9
Rigour	Methodological rigour	7	11
Rigour	Rigour (methods and approach)	8	5
Rigour	Methodological rigour	9	11
Scholarship	Scholarship, adding to body of knowledge	2	12
Scholarship	Scholarship and original synthesis	8	7
Scholarship	Advance body of knowledge	9	21
Utility	High value for users	2	9
Utility	Applicability	2	5
Utility	Utilization by end users	3	3
Utility	Impact on end Users	6	11
Utility	End user relevance	8	4
Utility	Professionally useful	9	5
Value for Money	Productivity	2	3
Value for Money	Valuable	2	2
Value for Money	Added value	4	6
Value for Money	Value for money	5	6
Value for Money	Value for money	6	3
	Well informed (engaged)	9	7
	Impact on quality of experience (human and other)	9	1
	Multidisciplinary and collaborative	9	1
	Impact on society	7	12
	Leadership in the field	7	11
	Authoritative	7	7
	Integrity	7	4
	Research income to subject area	7	2

---

[4] This sticky hexagon fell off the chart and was not available for voting.

Cluster	Characteristic	Workshop	Votes
	Theory based	7	0
	Accurate	7	0
	External funding	4	7
	Accessibility	4	2
	Deliverable output	5	4
	Quality at what level?	5	1
	Regional and national impact	5	1
	Holistic (judge team)	3	1
	Good project management	3	1
	Known in non-monolingual context	2	2
	Wealth creating	2	2
	Coherence	1	0
	Subject to transparent process at proposal stage	1	0
	Academic assessment	1	0
	Breadth of impact	6	13
	Number of disciples/students	6	1
	Inspirational	6	1
	Contribution to knowledge and understanding	8	13
	Puzzle driven	8	2

Table 2: Characteristics of Research Assessment Systems

Cluster	Characteristic	Workshop	Votes
Accountable	Accountability	6	1
Accountable	Accountability	8	4
Appropriate	Subject specific	2	9
Appropriate	Fit for purpose	2	2
Appropriate	Appropriate for discipline	5	9
Appropriate	Sensitive to changes in discipline boundaries	6	4
Appropriate	Sensitivity to discipline	7	13
Appropriate	Appropriate to expertise	8	9
Appropriate	Appropriateness of criteria for discipline	8	5
Appropriate	Appropriate method within discipline	9	0
Career Sensitive	Supportive of early career	7	4
Career Sensitive	Supporting careers	9	0
Career Sensitive	Sensitive to life changes	9	0
Comparable	Comparability across subjects	2	9
Comparable	Cross subject comparability	4	10
Comparable	Comparable and suitable	5	9
Comparable	Consistency across disciplines	6	2
Comparable	Parity of criteria within clusters	8	5
Comparable	Comparability of clusters	8	1
Comparable	Comparable across disciplines	9	7
Cost efficient	Cost effective	4	1
Cost efficient	Cost efficient	5	5
Cost efficient	Cost effective	6	3
Cost efficient	Efficiency	6	1
Cost efficient	Cost effective	7	1
Cost efficient	Efficient	9	2
Credible	Broadly acceptable	4	4
Credible	Credibility	5	9
Explicit	Clear rules	4	13

Cluster	Characteristic	Workshop	Votes
Explicit	Clear and acceptable criteria	5	12
Explicit	Explicit criteria	6	7
Fair	Fair and flexible	1	4
Fair	Fairness	2	9
Fair	Fairness	4	10
Fair	Fair	6	2
Fair	Fair	7	8
Fair	Fair reward	8	7
Fair	Fair to individual	8	4
Fair	Fair to institutions	8	0
Feedback	Meaningful feedback	4	3
Feedback	Good feedback	8	3
Feedback	Feedback	9	1
Flexible	Flexibility	3	5
Flexible	Flexibility across disciplines	6	10
Flexible	Flexibility across institutes	6	2
Flexible	Flexible timing	8	1
Flexible	Flexible across disciplines	9	8
Flexible	Responsive	9	1
Forward Looking	Forward looking	6	3
Forward Looking	Forward looking	9	0
Improve Research Base	Improve research base	4	4
Improve Research Base	Be able to advance field	5	7
No game playing	No gaming	1	0
No game playing	No scope for game playing	6	7
No game playing	Minimise game playing	7	6
Not Burdensome	Not burdensome	1	6
Not Burdensome	Not burdensome	2	13
Not Burdensome	Not onerous	4	3
Not Burdensome	Low burden	7	6
Not Burdensome	Not burdensome	8	5
Not Burdensome	Not burdensome	9	11
Objective	Objectivity	2	10
Objective	Objectivity	7	4
Objective	Objective	9	7
Output is Assessed	Dissemination of outcomes	1	3
Output is Assessed	Output is assessed	1	0
Output is Assessed	Biased towards output measures	6	7
Output is assessed	Assess output and not input	7	3
Peer Recognition	Peer review	1	12
Peer Recognition	Peer recognition	3	11
Peer Recognition	Peer review	4	8
Peer Recognition	Assess impact in peer group	5	5
Peer recognition	Peer recognition	7	1
Peer Recognition	Peer review	9	14
Quality not volume	Assess quality not volume	5	9
Quality not volume	Measure quality not volume	6	17
Rigour	Rigorous	2	9
Rigour	Rigour	3	8
Rigour	Rigour	4	8

7

Cluster	Characteristic	Workshop	Votes
Rigour	Rigour	9	3
Robust	Robust	1	11
Robust	Robustness	3	4
Transparent	Transparency	1	9
Transparent	Transparency	2	13
Transparent	Transparency	3	7
Transparent	Transparent process	4	13
Transparent	Transparent	5	8
Transparent	Transparent	6	9
Transparent	Transparent	7	10
Transparent	Transparent process	8	8
Transparent	Transparency of rules	8	4
Transparent	Transparent	9	18
	Consistent	1	3
	Inclusive	1	2
	Periodic metrics	1	0
	No spill over effects	1	0
	Dynamic	2	8
	Competence	2	5
	Legitimacy	2	4
	Consensus	2	4
	Match funding to results	2	0
	Qualitative	3	8
	Holistic	3	1
	No double counting	3	1
	Continuity	4	5
	Clear function	4	5
	Regional sensitivity	4	3
	Assessment impact on end users	5	6
	Citation analysis	5	4
	Praxis (integration of theory and action)	5	4
	Recognition of process as well as outputs (capacity building)	5	3
	Simplicity	5	0
	Clarity of outcome (in funding)	6	7
	Value different types of output	6	5
	External validation	6	4
	Timely notice of criteria	6	4
	Differentiate between fields according to research outputs	7	8
	Proper assessment of interdisciplinary research	7	6
	Independent judges	7	6
	Useful output	7	2
	Take into account end users	7	2
	Consequences reflect quality	7	0
	Contribution from international experts	8	6
	Reflexive	8	4
	Longevity of research	8	2
	Quantitative analysis where possible	9	8
	Systematic	9	4
	Multifaceted	9	3
	Respect different opinions	9	1

Cluster	Characteristic	Workshop	Votes
	Look at usefulness	9	1
	Joined up with other funders	9	1
	Longevity	9	0
	Accurate credit	9	0
	Retrospective	9	0

Table 3: Total Votes for Clustered Characteristics of High Quality Research

Cluster of Characteristic	Independent Mentions	Votes
Defining Research Agenda	8	99
Rigour	7	71
International Recognition	8	66
Original	8	61
Peer Recognition	5	50
Scholarship	3	40
Utility	5	37
High Quality Publication	4	32
Innovation	5	31
Academic Impact	4	26
Long Lasting	6	24
Value for Money	5	20
Creative	3	19
Breadth of Impact	1	13
Contribution to knowledge and understanding	1	13
Impact on Society	1	12
Leadership in the Field	1	11
Relevance	4	9
Research Culture	2	8
External Funding	1	7
Authoritative	1	7
Well informed (engaged)	1	7
Deliverable Output	1	4
Integrity	1	4
Collaborative	2	3
Known in non-monolingual context	1	2
Wealth creating	1	2
Puzzle driven	1	2
Research income to subject area	1	2
Good Project Management	1	1
Regional and National Impact	1	1
Impact on quality of experience (human and other)	1	1
Number of disciples/students	1	1
Quality at what level?	1	1
Multidisciplinary and collaborative	1	1
Inspirational	1	1
Accessibility	1	1
Holistic (judge team)	1	1

Table 4: Total Votes for Clustered Characteristics of Research Assessment Systems

Cluster of Characteristic	Independent Mentions	Votes
Transparent	9	98
Appropriate	6	51
Peer Recognition	6	50
Not Burdensome	6	44
Fair	6	42
Comparable	6	41
Explicit	3	30

Cluster of Characteristic	Independent Mentions	Votes
Rigour	4	27
Flexible	4	27
Quality not volume	2	26
Objective	3	21
Robust	2	15
Credible	2	13
Cost efficient	5	13
No scope for game playing	3	13
Improve Research Base	2	11
Output is Assessed	3	10
Qualitative	1	8
Differentiate between fields according to research outputs	1	8
Quantities analysis where possible	1	8
Dynamic	1	8
Clarity of Outcome (in funding)	1	7
Feedback	3	7
Independent Judges	1	6
Contribution from International Experts	1	6
Assessment impact on end users	1	6
Proper assessment of interdisciplinary research	1	6
Value different types of output	1	5
Competence	1	5
Clear Function	1	5
Accountable	2	5
Timely notice of criteria	1	4
Consensus	1	4
External validation	1	4
Continuity	1	4
Reflexive	1	4
Citation analysis	1	4
Legitimacy	1	4
Career Sensitive	3	4
Praxis (integration of theory and action)	1	4
Systematic	1	4
Consistent	1	3
Recognition of process as well as outputs (capacity building)	1	3
Multifaceted	1	3
Forward Looking	2	3
Output is assessed	1	3
Take into account end users	1	2
Inclusive	1	2
Regional sensitivity	1	2
Useful output	1	2
Longevity of Research	1	2
Joined up with other funders	1	1
No Double Counting	1	1
Peer recognition	1	1
Holistic	1	1
Respect different opinions	1	1
Look at usefulness	1	1

Annex C – Task 2 Assessing the Four Basic Systems

Workshop 1 Basic Systems Assessment

Algorithm	Expert Assessment	Historical Ratings	Self Assessment
Good Things Totally Objective Light Touch (entirely computerised?) Reduces game playing	**Good Things** Judgement is objective Best measure available Enable prospective judgement	**Good Things** Simple light-touch Often realistic assessment Audit based approach	**Good Things** Simplicity Recognises progress
Questions How to validate algorithms? How to recognise novelty How to ensure subject consistency	**Questions** How to recognise innovation? How to choose experts? How to deal with comparability between subjects?	**Questions** In combination with what other system? How to deal with criteria setting What time period is involved?	**Questions** How to recognise changes in patterns How to deal with innovation Subject or institution based
Questions How do you decide what's in the algorithms? Who decides what is in the algorithm? Does the same algorithm apply to all subjects?	**Questions** How do panels get chosen Should panels be international (entirely international that is)	**Questions** What is being assessed? Who would do the assessment?	**Questions** What would HEFCE used for this? Would the metrics ever change
Bad Things Not suitable for all subjects Not comparable for all subjects Open to game playing (because rules are very clear) Is having no peer review input acceptable?	**Bad Things** (They considered principally peer review) Biased panels – weak chair people Panels are too introspective Pressure on panels members – time (pressures of time) Pressure on panels outcome – outcome (pressure to produce positive assessments) Can panels be sufficiently comprehensive Time taken to do through assessment very long (and huge) volume of work Inconsistency between panels Not reproducible	**Bad Things** Internal Friction (within university) (between those assessing and the assessed) Credibility is lower (than peer review both internally and externally) Couldn't be used on its own Lack of consistency between returns (between universities)	**Bad Things** Sterile (10 years old or more) Changes since the last RAE not taken into account No leverage (difficult to put in directed change eg for more blue skies research) No credibility with time (ie credibility decreases with time as sterility increases)

Workshop 2 Basic Systems Assessment

Algorithm	Expert Assessment	Historical Ratings	Self Assessment
Good Things Objective if appropriately chosen for the discipline Need better markers Transparent	**Good Things** Competent, provided have real experts Legitimate Not burdensome (for reviewees)	**Good Things** Not Burdensome Less expensive to run Ability to plan, ie you know what your rating will be	**Good Things** Formative – opportunity to consider self Non-mechanistic Talking self responsibility, thinking about what you do Provides dialogue with assessors
Questions How are the measures chosen (needs to be discipline specific)? How many indicators? Need enough to be robust How are the indicators combined?	**Questions** How to define field? Who are the experts and how are they selected? How many experts?	**Questions** How far back Does rating include publications (makes rating even more out of date) How do you deal with emerging areas? How do you deal with declining areas?	**Questions** Who would pay for assessment, institution or funding council? Who sets the goals? It is easy to reach goals if you set them low How restrictive would assessment framework be?
Questions What do you want to count? (are they all relevant for all disciplines) How do measures relate to quality? Why do you use inputs to measure outputs?	**Questions** Who are the experts? Who decides? How can you be sure the experts follow the rules? How do you ensure coverage of multi/cross disciplinary areas? How do you avoid bias?	**Questions** What criteria and indicators would be used? Who makes the judgement? What period of ratings are used ie how far back?	**Questions** What criteria? How would they be compared? Is it retrospective or prospective? Over what time frame? What evidence is required?
Bad Things Double counting Quality blind – always using proxies, example of bad paper gaining many citations Strategically vulnerable Binds you to existing distributions Inappropriate for some disciplines (particularly Art History and Humanities)	**Bad Things** Coverage/Expertise Particularly Multidisciplinary[5] Lack of Multidisciplinary expertise Discounts innovation as 'big names' on panels Bias/Strategic manipulation Stature of developing area/identification of 'experts'	**Bad Things** Ossification/Inflexibility Purely retrospective, no account of future plans Acts against innovation and emerging areas Lack of credibility	**Bad Things** Credibility – both within and with out system Effort could be large (eg Teaching Quality Assessment) Subject boundaries (parochial) Not objective (no peer review) Game playing

---

[5] In feedback on the write up it was pointed out that the real problem may be assessing interdisciplinary research, rather than multidisciplinary research and the differences between the two need to be borne in mind.

Workshop 3 Basic Systems Assessment

Algorithm	Expert Assessment	Historical Ratings	Self Assessment
Good Things Cost Effective Objective Transparent	**Good Things** Sensitive to discipline Includes peer review Qualitative	**Good Things** Stability Good news for the Russell Group ie the rich stay rich Inexpensive	**Good Things** Sensitive to Discipline Sensitive to Institution Possibly Lighter Touch (but trade off with rigor of auditing, and may put more work on institution)
Questions How could 'Measures of Reputation' be reliable? How will you include non-standard research outputs? Will different subjects have appropriate algorithms?	**Questions** Who chooses 'experts'? What level of transparency is effective? Complete transparency may allow more game playing. Could also have transparency at different points eg elect experts and then those experts are trusted to devise rest of system Likely to be expensive, can it be afforded?	**Questions** How will you deal with significant change within an existing department/subject? How do emerging areas gain recognition and funding? How far back will ratings go? Which ratings will be used?	**Questions** How will it be validated in a cost effective manner? How much flexibility will institutions be allowed in structure of self-assessments?
Questions What algorithm? What weight to different metrics? How will variations across subjects be handled?	**Questions** What range of expertise? Will it include end users? Include experienced contract research staff? How will interdisciplinary staff be assessed? How to harmonise across panels?	**Questions** What time period? How to weigh up both absolute position and trend ie if an institution is getting better or worse?	**Questions** Who would you entrust with making the assessment within each institution? Are there common criteria? How would you police it?
Bad Things Not necessarily tuned to disciplines (so may disadvantages cross disciplinary research) Encourages inappropriate focus – ie focus on what can be easily measured Hidden bias although apparently transparent	**Bad Things** Self replicating (panels are 'usual suspects') Lack of consistency between panels Lack of transparency in weight given to metrics by panels (RAE2001 problem)	**Bad Things** Static Fails to motivate Discourages Collaboration – get 'class system' of departments Difficult to account for restructured groups/institutions	**Bad Things** Fails on transparency and peer recognition Fails to be comparison with correct reference group – comparing with others in institution not others in subject area Fails through lack of common confidence in system (eg Teaching Quality Assessment)

Workshop 4 Basic Systems Assessment

Algorithm	Expert Assessment	Historical Ratings	Self Assessment
Good Things Cheap Simple Transparent	**Good Things** Broad acceptability (within academe, just an extension of general peer review process) Inclusive Can be made international	**Good Things** Simplicity Light touch/low cost Institutional Ownership of process	**Good Things** (already part of existing system) Low cost Departmental ownership and trust of institutional management Compensates for narrow expertise on existing panels
Questions How to measure quality? How to ensure comparability? What would be the elements of the algorithm?	**Questions** Who chooses the experts? How do we seek to avoid prejudice? How to maintain transparency?	**Questions** How will changes be taken into account/monitored? How do you account for or invest in emerging disciplines? How do you motivate improvement?	**Questions** How would you police/validate? How do you penalise optimistic/inflated results? (or vice versa ie compensate pessimistic results) What evidence would be required?
Questions How do you prevent the process becoming the formulator, do you start behaving to get good metrics? How do you justify the metrics used? What evidence is there that they are right? How do you compare across disciplines?	**Questions** How do you widen and deepen expertise? How do you ensure cross-comparability across panels? How do you build in appeal mechanisms? How do you ensure appropriate criteria?	**Questions** How would you encourage new initiatives? How do you take account of changing individual performance and individuals moving institution? How would you ensure rewards are appropriate to current and projected performance?	**Questions** How do you assure validity? How do you assure reliability? How do you deal with self-deluded and manage feedback?
Bad Things Not cross field comparable Overly mechanistic, researchers end up chasing metrics Metrics are proxy measures Potential for spurious objectivity giving spurious respectability	**Bad Things** Ultimately subjective Not transparent Experts cannot cover expertise in any subject area, as areas are now too big 'Expert' is too narrowly defined, there is much expertise outside academe	**Bad Things** Inhibits change and leads to fossilisation Rewards inappropriately Perpetuates institutional and subject 'silos'	**Bad Things** Self deception Audit requirements/validation needed anyway so higher cost Tendency to preserve status quo

Workshop 5 Basic Systems Assessment

Algorithm	Expert Assessment	Historical Ratings[6]	Self Assessment
Good Things	**Good Things**	**Good Things**[6]	**Good Things**
Transparent	Specialist knowledge	Allows demonstration of track record	More responsive to disciplines
Clear and fine grained measurement	Community has experience of system – ie peer review	Builds on strength for the future	Encouragement of self-reflection
Cost-efficient	Confidence of Higher Education Institutions compared to alternatives	Could be combination of retrospective and prospective	Disciplinary autonomy
Could be anonymised (which addresses perceived biases)	Familiarity gives confidence	Can look at rate of improvement	More engagement with process
Questions	**Questions**	**Questions**	**Questions**
How would components of Algorithm be chosen and weighted?	How are reviewers selected? How many?	Clarification of term: is it a critical review of change or a measurement of what has happened before ('jewel in aspic')	How does it link with external review (a reality check)?
How would differential weightings be developed for different subjects?	How is multi/interdisciplinary work assessed?		By institution and or discipline?
How is quality discerned?	How are highly specialised areas dealt with equitably?		How measure against absolute standards?
Who makes the decisions?	How are decisions made seen to be transparent?		
Questions	**Questions**	**Questions**	**Questions**
How are metrics weighted within an algorithm?	Will all reviewers have enough time and capacity to undertake review? (particularly users and international)	How to overcome stereotyping in 'reputational' assessment?	Who assesses the self-assessors?
Who makes all these decisions?	How to assess interdisciplinary research fairly?	How do you overcome and remove inefficiencies?	Is unit of assessment individual or institution?
How do we recognise emerging areas			
Bad Things	**Bad Things**	**Bad Things**	**Bad Things**
Variable applicability to disciplines	Small sample of assessors	Conservative	Prone to self-serving bias
Many of suggested metrics are bad proxies for quality	Expensive	Doesn't allow for measuring improvement and development	Problem with comparability
Poor measure of long term impact	Opaque, the use of 'other minds'	Stifles developing subjects	
Unfair to young high quality researchers	Potential for bias	Looks at individual not collectivity	
Can cement status quo because of 'Arbiters of entry' to achievement of metrics	Tends to be conservative ie fails to reward innovation	Doesn't improve research culture, training and practice	
Would want to use very different metrics in different fields	Problems of comparability	Reinforces inequalities	
Spurious objectivity			

---

[6] In this workshop 'Historical Ratings' was mistakenly written as 'Historical Review' for the group tasked with looking at good points.

Workshop 6 Basic Systems Assessment

Algorithm	Expert Assessment	Historical Ratings	Self Assessment
Good Things Appropriate to some fields Objective Cheap Quick, which could allow for rolling review	**Good Things** Flexible Outputs are read Picks up subtleties, such as interdisciplinary research Responsive to complexity (eg different end users)	**Good Things** Low Cost Creates stability 100% transparent Reproducible Strengthens strong themes	**Good Things** Low cost to HEFCE Flexible Consumer input ie determined by researchers themselves
Questions How widely will algorithm be known (potential for game playing vs privileged knowledge)? How would weightings 'correct' for passing research fashion? How would reputation be measured (you scratch our backs, we'll scratch yours') How do we allow for differences between disciplines/research cultures? How would rolling review respond to changes?	**Questions** Who are and who appoints the experts, is there perpetuation of existing structures (old boy network)? How is agreement reached on subjective assessment? Who sets criteria and how? How is consistency maintained across different disciplines?	**Questions** Would trends be measured? How would improvement be encouraged? What happens about personnel changes? How would new departments and fields be measured? No incentive to change	**Questions** How would this relate to departments, would researchers do individual assessments? How do you judge accuracy of assessments? What criteria would individuals be given? What would be the units of assessment? Who would be the external assessors/validators?
Questions What do they mean by 'reputation based on surveys' and 'measures of financial sustainability'?	**Questions** How do you choose the experts? What are their interests and agendas – especially non-peer experts? Is it possible to blind assessments? How would you determine the 'right' experts?	**Questions** How far back? What criteria for assessing 'track record'? What are the assessment units, given changing landscapes? How do you identify a trajectory?	**Questions** Are criteria external for internal? What criteria for challenge to self-assessment? What penalties for consistent cheating? Who is 'self'?
Bad Things Biased against some disciplines – student numbers and research income aren't proxies for quality in all fields. Bibliometric measures can be inappropriate Open to abuse/misinterpretation (mutual citation) Disadvantages younger staff and staff at start of academic careers Quantity does not equal quality Research finance does not equal quality Surveys are arbitrary The system is ideologically flawed – the idea you can quantitatively assess quality	**Bad Things** Can be highly subjective Perpetuates 'Old boy' system Problematic if insufficient expertise in panels Could be driven by political agenda Time consuming and expensive Might not allow for interdisciplinary research to be assessed (if lack of expertise on panels) Potential for lack of transparency	**Bad Things** Not future looking Recursive – measuring a measure Still requires 'past performance' to be rated Lack of recognition of institutional starting point Not good at identifying trajectories Could create instability as academics flee to high ranking departments	**Bad Things** Game playing (high propensity to lying) Perceived illegitimacy Lack of broad perspective – too inward looking Just like Teaching Quality Assessment only worse Partial system – couldn't be used alone

Workshop 7 Basic Systems Assessment

Algorithm	Expert Assessment	Historical Ratings	Self Assessment
Good Things	**Good Things**	**Good Things**	**Good Things**
Low Cost	Assesses innovation	Reduces burden of assessment on some	Highlights strengths
Time efficient	Flexible and allows judgment	Top departments will remain at the top	Sensitive to local conditions, missions and disciplines
Transparent		Transparent	
		Minimise game playing	
Questions	**Questions**	**Questions**	**Questions**
What freedom would disciplines have for varying weighting across metrics?	How would you minimise bias/self interest?	How do you get incentives?	How is comparability ensured?
Why use only one method?	How would you minimise workload and cost?	How do you prevent inertia?	Who polices HEIs' claims?
How would you ensure data was credible?	How would you select experts – who? How could this be made representative?	How do you develop new areas or units?	How do you prevent game playing?
How does this method ensure dynamism/innovation in the system?	How would you ensure transparency of process?		
Questions	**Questions**	**Questions**	**Questions**
How should formula be decided upon?	Composition of panels – how decided?	How do you assess young researchers and research groups?	How do you validate assessment?
Frequency?	How will collaboration/interdisciplinary research be handled?	Who gets the rating, group or individual?	At what level does the assessment occur (individual or institution)?
How to assess long vs short term impact?	How will personal excellence be rewarded?	How do you allow for change?	
Bad Things	**Bad Things**	**Bad Things**	**Bad Things**
Crude assessment	Subjective assessments (not transparent)	Unresponsive to change	Subjective
More tailored to 'sciences' then 'humanities/arts'	Variation between panels	Inhibits aspirations	What are criteria
Ignores users of research	Possible self interest	Breeds complacency	Honesty
		Can't cope with potential – individuals and areas	Comparability
		How can new departments be handled?	

Workshop 8 Basic Systems Assessment

Algorithm	Expert Assessment	Historical Ratings	Self Assessment
Good Things Objective Simple Transparent Links to allocation can be clear	**Good Things** Reputational Provides knowledge and understanding Better for international aspect (by use of international peers) Involvement of users and industry Peer review is accepted Able to understand	**Good Things** Stability Good for status quo Ease of operation Looks at what has been achieved rather than speculation on future	**Good Things** Gives individual responsibility Brings out positive thinking Cheaper and simpler Allows reflection on change in departments etc
Questions How would you reconcile large departments and small departments Who designs the algorithm Weightings	**Questions** Who nominates experts? Who selects experts? Who chairs the panels? Period of assessment (future gazing) Is it transparent? Size of panel (could be larger)	**Questions** What elements should we use? What length of history? Can it work as a predictor?	**Questions** How would it be validated? How would you prevent grade inflation? What is the right level of assessment (individual, group, department, institution)
Questions How would discipline based 'differences' be taken into account? Is assessment based on individuals or total submission? Eg survey Who decides the framework?	**Questions** Who is the expert/who decides? How to deal with small or highly specialised subject areas?	**Questions** What is a track record? How far back? Intervening changes – inst/individuals How do you cope with interdisciplinary research?	**Questions** How do you decide on when to drill down? What is the self? How do you deal with decline in a unit? Appeals and litigation How do you deal with competitive aspect (Teaching Quality Assessment works as isn't competitive) How do you involve external input at institutional level?
Bad Things Weighting criteria need identification Not applicable to all disciplines eg arts, humanities, music Not sensitive to quality of content Open to manipulation Not 'objective'	**Bad Things** Small units of assessment have particular room for bias Resource intensive Lack of transparency through process - accountability	**Bad Things** Coping with change Conservative/complacent (stifles innovation) Assumes adequacy of previous system Perpetuates inequalities/inequities	**Bad Things** Systematic differences between institutions Game playing Belief in equivalence of judgement Risks institutional authoritianship Burden on nervous institution Shifts from subject to institution

Workshop 9 Basic Systems Assessment

Algorithm	Expert Assessment	Historical Ratings	Self Assessment
Good Things Objectively measurable Easy and automatic Non-burdensome Easily checked Easy comparability Easily used to apply funding as output is a number Easily adjusted and tweaked Credibility Flexible allowing different algorithms for different fields	**Good Things** Flexible/sensitive Long standing and widely used Allows prospective assessment Anchors UK research in international scene Allows incorporation of 'non-measurable' quality to be included	**Good Things** Stability helps planning Light touch and reduces burden Transparent	**Good Things** Researchers know their work best Researchers know where their research lies Encourages people to be reflective
Questions What measures should be used? Who decides what should be measured? Over what period? How would particular metrics be weighted?	**Questions** Who are the experts? How are they chosen? Are they representative?	**Questions** Is it exclusory? Is it accurate? Does it stifle creativity? What is the effect on morale? What about peer review?	**Questions** Would it have credibility?
Questions Type/mix/weighting? How do you get comparability between units of assessment (eg science and arts)?	**Questions** What data is used? (And what is read?) Who are the panel? Do experts have their own agendas?	**Questions** How far back? How to avoid bad points?	**Questions** How do you compare universities? How would you ensure international credibility?
Bad Things Spurious rigour Variability vis-à-vis disciplines in income and citation practices Validity of metrics Relies on assumptions eg correlation between metric and quality Algorithms don't exist in all subjects You can't quantify all aspects of quality Easily modified by HEFCE post hoc Entirely retrospective	**Bad Things** Possibly inconsistent between panels and subjects Matching experts to subject groups and areas Includes unreliable prospective assessment Hidden criteria may be used Small fields have bias/self interest problems Not transparent Domination of panel by few members	**Bad Things** No recognition of change – favours status quo Perpetuates elitism No value to forward thinking Encourages people to rest on their laurels No recognition of young disciplines and researchers Assessors influenced by submission	**Bad Things** Not objective, so low credibility Selective audit Has not worked in other sectors and areas (eg Teaching Quality Assessment) Halo effect with inflated claims Institutional pressure to overrate departments No international perspective Burdensome if done properly

Annex D – Hybrid Systems and their Implementation and Implications

The write up presented in this section are based primarily on transcriptions of the flipcharts prepared by the participants, supplemented by notes on discussions.

Hybrid System 1: Based on Self Assessment

Based on self assessment as they felt that was best basic system.

Added elements from other 3 systems.

Basic idea is to trust universities, institutions and academics to come up with an assessment of themselves.

Assessments would be made at a set time every 3, 4 or 5 years.

20% of assessment submitted would be reviewed by a panel of experts (this review could be triggered in three ways).

If there was a big difference between the assessment based on metrics (eg research income) and self assessment.

If there was a big difference between historic rating and self assessment.

A random sample of other submissions.

This review would be by peers, much along the lines of current RAE assessment and would contribute the idea of robustness, the idea of reviewing the research and the notion of peer review.

Discussion

Would outputs/research publications be read and assessed? Only if a review of the submission was triggered.

Would the self-assessment be published? There hadn't been discussion of this.

Would criteria for self-assessment be set across the review? Could vary by subject.

What level would self-assessment be done at? Same size as at present ie by department. Alternatively, could do subject areas by QR funding groups which should help problems of assessing multidisciplinary research.

Similar selection of academics submitted previously – ie don't have to submit all. But there were worries that this could undermine robustness.

Implementation

Agree metrics and disseminate.

Agree Units of Assessment.

Agree mechanistic framework for assessment.

Agree criteria for each Unit of Assessment, wouldn't need to be the same.

Agree framework for auditing of self assessments.

All of this needs to be done early in the implementation process.

There should be fewer units of assessment.

Discussion

This new self assessment RAE could alternate with a full RAE to prevent grade drift and keep system calibrated.

What could go wrong?	Hoped for changes in UK research
Wide scale dishonesty, there was discussion over what penalty for 'wrong' self-assessment should be.	Simpler to operate, so less wasted resource, both in preparing submissions and reviewing them.
Challenge to process: arising from a feeling that either it is unfair if you are audited or unfair if you are not (maybe most audited departments get given higher rating).	Easier to get 'Experts' to do reviewing in audits, because there would be less work. Could also get better experts. International experts should be easier to recruit, as they would only have to review audited institutions claiming 'international' level.
Danger of departments that are getting weaker being missed, as departments 'staying the same' would not attract attention.	Maintains benefits of RAE in terms of quality improvement.
Process brought into disrepute. RAE is generally accepted, would new system be?	Less games playing as there might be less obsession with the system than there is with the RAE.
Danger of grade drift.	

Hybrid System 2: Based on Expert Review

Similar to current system, but tweaked.

Longer period between reviews (ie > 5 years).

Could produce departmental 'research profile' initially eg 30% of research at international level, 20% national and 50% lower. Then use an algorithm to convert this into a rating, so there is more than one way to get each rating.

Panel selection should include international members from the start, to ensure they are up to speed. Universities should also be allowed to nominate panel members (but probably not members of their own institution).

Panel members should be seconded full time.

Fewer Units of Assessment which should help make process more consistent, this would produce more subpanels.

Panel determines what data will be required in advance and this is then subject to external challenge (could be input from previous panel, foresight panels) but final say is with current panel. Different panels could request different data.

Could look to include 'success rate' of research applications as data – closer to current situation than current 'research income'.

Panel chairs should be 'lay members' may be judges or recorders.

Discussion

Should funding go through these subpanels or should they simply feed into their panels?

Should all staff be assessed?

Should assessment be at the level of individuals or groups?

Would it be hard to run if each panel could choose their own data requirements?

Shouldn't be too bad with reduced number of panels, but could have pick'n'mix list of criteria which panels could pick from to limit total number of criteria.

Could you address innovation? By peer review and by using current/pending research grant awards rather than research income. However there might be problems with auditing this data.

Implementation	What could go wrong?	Hoped for changes in UK research
Select data for exercise and decide on number of Units of Assessment.	Human error – risk of international panel members dropping out/accepting in first place, similar problems with lay chairs.	Greater transparency.
Capture wisdom from RAE2001 exercise, maybe old panels could create pick list of assessment criteria.	Criteria and new data – use pick list to try and prevent this becoming too unwieldy.	Better result that continues high regard in which RAE was held.
Nominate chairs for panels – HEFCE for fund and arrange secondments.	Software for electronic submission needs to be bug free – use Microsoft preferably, to get less bugs	Shallower learning curve for whole sector as system similar to previous RAE.
Panel selection (Universities get to nominate).	Panel could resign.	
Panels meet to set criteria.		
External Challenge to criteria.		
Panels determine pick list of data to be provided (HEFCE in consultation with working panels).		
Decide on weighting/rating scale and number of staff to be included.		
Higher Education Institutions to produce data (need to look at whether there could be Higher Education Statistics Agency and RAE consistency.		
How are outputs to be submitted – preferably use electronic system.		

Hybrid System 3: Based on Expert Review

Starting point was Expert Review (most qualities they liked).

Could incorporate algorithms for comparability between panels – but can this work from science to humanities.

What metrics to use – for science use citation and impact factor, humanities – research council funding, as application includes element of peer review. Number of PhD students (and post docs) also a good measure for all.

Need to check relevance of metrics with research users – Learned Societies (this could differ between UoAs).

Review panel could use external referees or second extra expertise when necessary.

Could also use more subpanels – these could also call on expert reviews to give extra expertise for multi discipline work.

Build in structured self-assessment to provide context for assessment – recent appointments, evidence of esteem and infrastructure.

Could assessment be more frequent? Maybe once a year? Could you make more often but less heavy such that information was held centrally and simply continually updated (how accurate is centrally held data?). Also might help to avoid saw tooth effect on staff appointments and publications.

More frequent assessments could help new departments – as they need to be assessed early, or could have possibility for interim assessments. Panels could determine this as they should know relevance of each output.

Also want to increase scope of what counts as research output, this is a particular problem in humanities eg exhibitions. Panels could determine this as they should know relevance of each output.

Could also allow all publications to be shown with most important ones highlighted, this gets over issues of different number of publications in different areas.

Discussion

Why use Learned Societies to define users? Only a first suggestion.

Implementation	What could go wrong?	Hoped for changes in UK research
(Of slightly modified plans)	Algorithms can lead to game playing.	Smoothing of resource – aids planning.
Consultation with academics and non-academics (nationally and internationally) to define criteria and algorithm.	Grade inflation.	Inclusive and less game playing and linked, allows development of proper research culture – not dog eat dog.
Elect panels by academic community – add users and cooptees.	Different measures may cancel each other out, so get no spread of excellence.	Ensure research capacity is activated wherever it is found.
5 year notice of changes (at least).	More burdensome.	Better career structure and development.
Earlier software development.		

Hybrid System 4: Based on Expert Review

Supported by appropriate metrics (including value for money and input-output).

Need more than numerical input and output measures – ie need to relate output to money in. Panels should be doing this but possibly not in entirely mechanistic way.

Also need to take account of time devoted by staff to research ie need to count other responsibilities.

Prospective planning takes form of bidding on strategic plan (a form of self assessment) and they judge this in five years hence. There is an implication that 'Quality Related' money will go to bidders.

Historical data part of on going assessment to allow monitoring of incremental change between assessments.

System would take greater account of personal career progression eg addresses institutional sexism.

OR

Review dual support by increasing research council funds.

Discussion

The equal opportunities point is very important and also need to take account of other issues that draw people away from research from time to time.

Who would assess strategic plan? Could be done by panels or other experts pulled in for expertise.

Implementation	What could go wrong?	Hoped for changes in UK research
Set up group to investigate exploring value added and how to assess it.	Some measures are two edged.	More forward looking process.
Introduce 5 year business plan as key component.	Could become over complex.	Better planning of research development.
Move from grades to marks – should summing be completely transparent, probably not.	Reduces management flexibility if assessing strategies.	Research should become better value for money if this is assessed.
Introduce incremental change process (adjust core volume, trigger reviews if threshold is reached).	Game playing, but this will always be the case.	More responsive to change.
Include all staff and take account of career situations.	Head of Department will put pressure on staff (hard to avoid).	Improved personnel career development.

Hybrid System 5: Based on Expert Review

Aims

Need system that is more responsive to change.

Need some aspect of values for money.

Need differing mechanisms for differing subjects/panels.

How?

Need 'extended' peer review combining self-nominated referees and panel.

Variable timing of assessments.

Use Metrics to trigger a new assessment, if there is a change this could trigger peer review assessment. This would allow downward turns to trigger review.

Input from users of research – self-nomination of users and should be subject variable.

Whatever system is used all people should be submitted this is important.

How could grading be translated into funding?

Need to look at return on investment over the previous cycle.

Separate grading from return on investment.

Remove step changes.

Could award points for each individual and then combine these.

Should be no minimum number so no minimum size.

Implementation	What could go wrong?	Hoped for changes in UK research
Advanced notice of transparent criteria with peer input.	Poorly designed metrics.	More responsive to change.
Determination of how funding would follow this.	Poorly selected panel (and extra panel experts).	Broader input to decisions.
Establish metrics mechanisms in a subject specific manner.	Excessive cost and time.	Equal opportunities improved (as all submitting) both individual and institution.
Panel composition, including metrics experts and funding experts such as HEFCE, Research Councils, Charities etc.	Frequency of peer review – too long or too short.	More progressive change (rather than step changes) this has implication of permitting better planning.
Use average of individual score of ALL staff.		
Scores for broad areas for comparisons.		

Hybrid System 6: Based on Expert Review

Panel of reviewers has to be main contributor.

Each panel should have 'external' member from related discipline to give comparability and useful input.

How could fields be better judged by sub panels in a UoA?

All submissions need to be assessed, often reputation leads to nod through.

Need panels determined by extended stakeholder groups: research councils, specialist associations, international experts.

Algorithms need a value added set of criteria as well as value for money need to take into account development of new areas and building infrastructure.

Need to give nurturing institutions credit for subsequent success of high flyers.

Discussion

Is it practical to assess all submissions? Ask all researchers to rank their returns and specify that everyone's best paper should be read

Implementation	What could go wrong?	Hoped for changes in UK research
Need to do consultation to determine panels that is not linked to current structures of academia. Appoint panels: panel members, expert reviewers, international reviewers (non-UK and non-US), and stakeholders beyond existing structures. Panels collect information. All of the submissions go out to external reviewers for rating and then go to panel, panels job is also to add in value for money issues and take account of algorithms.		

Discussion

Degree of socialisation before reading submissions is large, this system might gain on degree of expertise but loose on area of process standardisation.

Hybrid System 7: Based on Expert Review

How do you select the experts?

Need greater transparency in how the panel members are selected.

International experts, need to be paid enough, and need to be fully engaged in system. In RAE2001 they performed badly.

Some disciplines have specific problems with international experts eg French French historians have very different views to British and American French historians.

Other Points

Read all submissions, there at least needs to be consistency across panels about how much is read.

Metrics could be used but should only inform and not determine rating. Metrics need to be sensitive to subject and could be determined by panel. Could include % of graduate completions, research money etc.

Institutions should put together their own submission.

Only exclude staff who solely teach. Should be submitting 75 – 100% of staff.

Need a better system for new researchers, RAE2001 was heavily biased towards established researchers, eg by suggesting four papers.

Need some way to provide money to high rating departments– need to make sure that it gets to the researcher.

Use a continuous rating scale rather than step changes between grades.

More on metrics – look at different ones for different subject areas (% post grad completions, funding etc), chosen by the expert panel.

Discussion

Discussion of whether all submissions need to be read: Should there be different numbers of submissions for subjects, depending on value to breadth. Is re-reading necessary for fields were peer reviewed publication is well accepted – why re-read a Nature paper? But RAE allows you to take account of such factors as career development, and not all fields look to Nature papers as a gold standard.

Implementation	What could go wrong?	Hoped for changes in Scottish research
Establish pool of experts (determine criteria for selecting them, the number of panels and experts and allow opportunity for objections). Identify appropriate outputs and metrics (a core for all panels plus discipline specific). Determine assessment criteria and weighting for combining criteria. Collect identified outputs and metrics. Apply assessment criteria. Appropriate allocation of funds, need to take into account both individuals and departmental excellence.	Inconsistency across panels. Problem recruiting sufficient number of committed experts. Negative impact on teaching staff if they are excluded from doing research – much of this is cultural perception.	High quality researchers get funding regardless of average department performance. Better support and career structure for young researchers. Good support of emerging areas of research. Funding spread over broader base with enhanced funding for excellent areas.

Hybrid System 8: Based on Expert Review

Need methods of measuring external funding streams and other sorts of support and contributions in kind to include in metrics. For example £100K for a film project might never enter the universities books.

Need to balance pitfalls of not including all researchers and alternative pitfalls of submitting everyone – could to assess teaching and research at the same time and split FTEs.

Management of research careers is not currently well dealt with, one problem is not including research outputs of researchers (and Contract Research Staff) that are not currently captured within 'active researchers' set. Most of this could probably be captured qualitatively and could be done by panel assessing institutions structures and plans. Panel could be told to place more emphasis on this aspect.

Process for adjusting boundaries of UoAs as they don't necessarily match onto institutional structures. Currently people UoAs are built up from researchers from different departments.

Need sensitivity to discipline in detail of assessment process.

Need to address issues for inconsistency across panels – maybe chairs of panels should be from another discipline.

Include assessment of dissemination of outputs to (professional and public).

Discussion

Don't label UoAs, use only number and description.

Need experts that can judge anything submitted to the panels, this is hard at present and would be harder if few panels. Would panels be larger?

Implementation	What could go wrong?	Hoped for changes in Scottish research
Institutions should identify their key strengths.	Top up fees attracting researchers to England because of better resources.	International aspiration, should aim to be players internationally.
Assess institutions on internationally comparable benchmarks.	Mis-use of RAE funding by institution ie not given to good research groups/department.	Improved commercialisation and exploitation.
Develop career sensitive assessment measures, ie don't compare professors and postdocs, to allow inclusion of all staff in exercise.	Worry of 7-8 year gap, loss of institutional memory and unfair to judge institutions against plans from 8 years ago.	Institutions should focus on key strengths (including latent strengths).
Allow qualitative commentary on individuals and groups in RA 2?.		
Panels determine appropriate metrics and consult community.		

Discussion

Interesting that both groups have suggested that panels should be appointed at the beginning which gives them flexibility and power.

Problem of twin cultures of teaching and research – every person should be included in an assessment then this might be less of a problem, could make people 40% researcher and 60% teacher and then assess against total FTEs for teaching and research.

Hybrid System 9: Based on Expert Review

Considered general issues affecting research in the UK.

Decided they liked the idea of systematic review to bring struggling departments up and allow researchers to flourish.

Thought expert review was important as the central part of the system.

International expertise and wider expert needs to be brought in.

Metrics might be useful in some areas, because of greater precision and sensitivity.

Self-assessment has valuable aspects as it is about self worth – but who would determine extent?

Who chooses your panel? (In RAE 2001 some people were moved between subject areas by the panel).

Rules should be set and not changed – rules were changed during RAE 2001 process.

Need to cap highest grade that could be achieved if less than 75% of staff submitted.

Should be common and transparent process across panels and this should be set at the start.

Feeling that sampling was a recipe for lack of trust and bad faith.

Need appeal system.

There are also issues of old boy network, RAE chairs should not be 'inherited'.

Sensitivity to regional issues – there should be no special pleading, however, they felt that they were benefiting from Welsh assembly issues.

More sensitivity to 'joint submissions' - between subject areas and between institutions.

Could widen or narrow grade spectrum.

Clear definition of who has submitted, should category C and D[8] be abolished for submission?

Need clear rules for transfer system and how this relates to 'value added' – 'Centre Forward Syndrome'.

More specific feedback from panels.

Discussion

Didn't the last RAE recognise the transfer system? Yes but it is more a question of whether the use of the system is appropriate.

Should be recognition for the cradling department when they have developed career of subsequent high flyers.

There will be some reward for the cradling department, as they will have published there.

There will also be movement in the other direction from high level departments.

Issue of returning everyone is a key issue, also think that a key issue to look at is joint assessment of research and teaching? Part of this is an issue of what the system is for, it is now being used for all sorts of things. They should be kept separate, but there is an issue of how teaching money should be allocated.

Implementation	What could go wrong?	Hoped for changes in UK research
Early decision on date and timetable.	Last minute changes to rules or systems.	Gearing of grading to funding model in Wales.
Clear detailed guidelines and subject criteria.	Software that doesn't work.	Increased collaboration and more joint submissions.
Wider consultation on panel members (chairs should be elected by panel).	Inappropriate influencing of panel members.	Comparability of funding work with rest of UK
Software that works.	Financial outcome fails to reflect increase in grades.	
Training for participants.	System overwhelmed by number of appeals. Would have to determine if appeals could lead to drop in grading.	
Determine and publicise structure of financial settlement of Quality Related funding.		

---

8 Category C staff are other individuals (ie not submitted elsewhere) active in research in the department as independent investigators. Category D are individuals active in research in the department as independent investigators.

Hybrid System 10: Based on Expert Review, specifically RAE2001

Started from the RAE 2001 system and tweaked it to deal with concerns.

Panels

Number of panels – a feeling that this number could come down, splitting up subjects doesn't add anything. Maybe look to do size of panels by volume of work that was submitted to them.

Uniform processes across panels for transparency.

With fewer panels it would be very important to select panel members for breadth of knowledge.

Broader panels should find it easier to assess cross disciplinary research.

International assessors – this was the most opaque part of the process and it was felt that it should be made similar to the very open level of the rest of the system. Bottom line was either to radically improve the international system or scrap it.

Who to Judge?

Category C is open to far too much gamesmanship so this area of returns should be scrapped. But need to look at how this could be adjusted for NHS and clinical research, feeling was they should either be put in RA 2[7] or put in RA 5[7] narrative.

There are still some anomalies that were addressed at the 11th hour last time round. Some professional teaching staff have contracts excluding research, but research fellows who only do research aren't eligible. Maybe only staff involved in both research and teaching should be eligible.

RA 2,3a, 3b, 4[7] where considered okay.

RA 5[7] should be either used by panels or reduced to one page summary – use it or loose it.

RA 6[7] should be used consistently.

Also noted improved method for submitting and assessing cross institution collaboration. Last time they seemed to be making up rules as they went along, and very little feedback was provided.

Discussion

What about panel numbers? A specific example: engineering reduce to 1. More generally panel size should reflect discipline size.

Need consistent level of feedback between panels – *some panels provided good feedback others did not, it was felt this issue was tinged by the threat of litigation.*

Implementation	What could go wrong?	Hoped for changes in UK research
Announcement of date – not before 2008.	Miss critical dates.	Further improvement in UK research base.
Define rules as early as possible – by end 2003.	Concern about litigation – funding councils should be willing to take more litigational risk, or use clauses for institutions signing up to RAE to preclude litigation. This should allow more feedback.	With broader panels should be able to cope with more interdisciplinary work and hence should allow more innovative research.
Appointment panel members for units of assessment by end of Feb 2003.		
Panels consult on criteria, completed by mid 2004.	Comparability with rest of UK might be lost.	More closely aligns RAE with research councils and govt policy encouraging interdisciplinary work.
Definition of funding implications by end of year 2004.	Lack of clarity on any aspects.	
	Changing of rules or/and outcomes.	

Hybrid System 11: Based on Expert Review

Need to decide what the system is for, is it for kite-marking, or funding or what?

Does the system need to be comparable across disciplines if only number of returns determines size of pot for each subject?

Expert Review is probably the best of a bad lot, but there was potential for huge savings by using metrics in areas where it is appropriate.

Base system of expert review plus metrics with details devolved to panels, exact method declared at the start of the assessment cycle.

Shape of panels should be broader scope fewer panels (although not in medicine) – also a worry that interdisciplinary work was not being valued.

Need more feedback in reviewing process and possibility to respond to criticism, because peer review is flawed needs checks and balances.

How do you define the panel? Could they be constituted in different ways and how could their expertise be broadened – eg could they adjudicate on external reviewers comments rather than reviewing work themselves.

Maybe the system should be a continuous system of marks or a distribution of scores for the members of the department.

Concern that you have a 5* department where you have only declared one of the ten academics – is that better than a department with eight 4 rated people?

Implementation	What could go wrong?	Hoped for changes in UK research
Agree system early and include funding rules that will be applied	If system is known it is easier to 'play' the system	
	Fixed size cake – not funding absolute excellence	
Set out general rules	Implementation left too late	
Decide on panel members	Focus on the exercise, not the research	
Panel to consult subject communities	Comparability between disciplines is lost	
Decide and announce panel rules	Translation and dissemination is lost	
Start collecting information		

Discussion

If panels were agreed earlier then they could engage in dialogue with community to select criteria.

How and who are selected for panels?

Need to recognise development of research strategy and leadership, and some sort of prospective assessment eg for departments that can be shown to be up and coming.

Hybrid System 12: Based on NA, assumed there were elements of all.

No overall system picked because group could not decide.

A number of points for improvement on current system and issues raised.

Should have evolutionary development of current system.

Need to look at networks and collaborations, and how they can be encouraged and assessed.

Should reward growth in quality.

Considered cluster of international universities that were boxed off into separate system, but rejected this.

Star rated groups or areas should be able to attract specific funding

Want to include more diversity of input in the way of Experts – CBI, DTI, users etc.

There was a big debate about whether everyone should be submitted or not. This was not resolved.

Could also use a matrix model in which disciplines are clustered, looking at depth within discipline and context of work ie breadth.

Try to avoid encouraging transfers that would simply have the effect of rewarding the new host institution.

Regroup subjects into clusters and have super and sub-panels.

Discussion

Labelling of people as teaching or research – this was not a good thing and would impact on morale.

The bigger you are the easier it is to sustain excellence.

Sustaining research and keeping it dynamic are probably always in conflict.

Definition of criteria is improving which is important.

If the funding council gets it wrong it have terrible impact on entire research base.

Hybrid System 13: Based on Self Assessment and Expert Review Hybrid

System of core and competitive funding.

Hybrid of Expert Review and Self Assessment.

Expectation of rolling process of self assessment with periodic moderation and validation by experts, maybe every five years.

Permanent developmental role for expert review panel.

Bid for competitive funding using self assessment.

Assessment should contained historical narrative and prospective review.

Panel provides framework for self assessment and develops concept of what excellence is and how it is measured.

Statement produced in transparent manner.

Categories for assessment include culture, outputs, infrastructure, stakeholders, publications. Have a system that rates each category, maybe score 1-5 for each area so have more discriminating scale and could look at specific area ratings for certain funding allocations.

A set of assessment methodologies and tools developed by panel.

Have different institutions/subjects being moderated/validated at different times.

Make units of assessment much broader: maybe social science one year, physical science next etc.

Discussion

Assessment systems was easy, but it was hard to work out mechanism when it came to allocating money.

Novelty of using self assessment is the prospective aspect of assessment – but this is hard when you are trying to anticipate what funding you think you will get – do you ask for the world or make very conservative plans? Could look at drawing up three scenarios of funding.

Implementation	What could go wrong?	Hoped for changes in UK research
Determine number, remit and contribution of UoAs/expert panels (need consultation and decision).	Demotivation.	Increased motivation.
Define relationship between role of panels and role of institutions.	Burdensome for institutions and panels.	Increased fairness in funding diverse activities and missions
Panels to define mechanisms for and objectives of self-assessment process.	Comparability.	Positive impact on strategic management of research.
Panels to decide rolling schedule for assessment.	Inaccuracy – problems of exaggeration.	Improved sustainability and flexibility.
Decide core funding allocation and distribution.		Consistency of funding environment, because of permanent steering group.

Hybrid System 14: Expert Review

Came up with framework in which assessment system would operate.

Felt that different cognate subjects areas should be able to generate their own research assessment systems, they feel that assessment method is currently science dominated.

There should be a few large UoAs which could determine their methods of assessment, based on research councils, and then subpanels within these.

Peer assessment and self assessment should be part of method for all systems but final method should be decided by over panel.

There would be sub panels for subject areas with some autonomy of system for sub panels.

Positive change must be recognised and this has an aspect of value for money.

There should be no bar on entry to any research assessment process.

Six year timescale seemed reasonable.

Need a grade scale that has less jumps in it.

Funding should start further down the grade scale (3a seems high point to start), if this isn't changed there will be a de facto separation of teaching and research.

Need to retain internationally competitive departments, but the notion of international excellence may not be relevant for all fields.

Implementation	What could go wrong?	Hoped for changes in UK research
Each individual staff member is clear as to which UoA they would be in as they would correspond to Research Councils.		Not all staff to be returned allows smaller institutions with less track record to do well.
Panel members, should include people who were current or recent members of RC panels. Could have a core panel plus – replicating EPSRC college system.		
Need to take pressure off panel members – could try to look at more metrics.		
Metrics produced 'centrally' by HEFCE/research councils from existing data and use of appropriate technologies. So institutions aren't recollecting statistics.		
Need good software.		

Hybrid System 15: Based on Expert Review

Something akin to current system.

There should be a longer time scale with triggering of intermediate review at 4-5 years (would allow for new entrants).

Reviewers should be peers, end users and international and make sure that the panel churns 100% every ten years.

Need a more subtle scale with more points, this could be splitting 5* ie high five star or low high star.

Improved scaling should also be reflected in individuals - current regional, national, international is very crude. Could give score for each individual and sum them for UoA.

There should be capping of maximum grade if not everyone is submitted.

Reduce number of UoAs and align with Research Councils.

Within UoAs assessment system could be modified to make it more appropriate to subject, some panels might have more end users.

Use a weighted algorithm, to take account of per capita income and total research funding in discipline.

Include aspects of transfer fees and compensation for nurturing institutions.

Some panels may have no peers.

Discussion

How would the weighted algorithms work? Need to take account of differences within subjects eg theoretical vs practical. There would be option of the panel to use or not use research income as measure.

Implementation

General units of assessment based on Research Councils.

Steering committees for UoAs in year one, they should do benchmarking for each of the UoA this should involve Prof Societies, End users, other major funders, and departments (not university administration).

Steering committees determine panel composition principles and define appropriate algorithms if any.

Over arching UoAs panels would identify sub panels (fewer than current number) could look to make them less subject based, aim to encourage interdisciplinary research.

Publish and disseminate criteria for award and revised funding model.

Choose panel - based on appropriate criteria and should have transparent appointment process, including interviews.

First assessment.

Trigger for mid term assessment is either volunteering (who are prescreened) or by major change in circumstances.

Discussion

How does this map onto department? Individuals could submit to their select panel - not all in department would have to submit to same panel.

What could go wrong?	Hoped for changes in UK research
Institutional misallocation of resources.	Less volatility in funding.
Disciplinary boundaries preserved.	Longer research horizon.
HEFCE misallocates resources.	More interdisciplinary research.
Abuse of mid term trigger (use pre screen to stop this).	Better fit between quality and reward.
If it is based on areas of Research Councils is there a problem of Research Council double impact which might compromise dual support system.	More transparency.

Hybrid System 16: Based on Expert Review

Crucial issue was to decide panel composition including role of peer and non-peer groupings.

Composition must be sensitive to UoAs and would be negotiated between representatives of UoAs and HEFCE. This consultation should include - national bodies (eg Learned Societies, subject associations, professional bodies, international organizations and departments).

General favoring of peers over non-peers.

8 to 10 year cycle with mid point.

Main review heavily dependent on expert review – and to produce more feedback to institutions possibly including weaknesses and strengths and targets to lead to upgrading. Targets could be negotiated in dialogue with panels.

At mid point ideally lighter touch reassessment on basis of meeting targets.

Need to have stick to make sure good departments do not get complacent so would have to rejustify their rating.

Mid point would probably have more self-assessment.

Did not look at whether assessment of everyone occurred.

Subject specific qualitative evaluation of narrative submissions describing non-written outputs eg curatorship.

Should be equivalence of output for example one book could equal ten journal articles.

Assessees can propose additional outputs through dialog with panel.

Explicitly not counting income generation.

Discussion

There needs to be subject sensitivity about whether research income is used.

How would you deal with new departments and subject areas? Not directly addressed, but could be looked at as part of intermediate assessment.

Is there mileage in having a special track of assessment for new departments keeping them apart from established departments?

How do you prevent mid point becoming a new RAE? This is a big issue - need to look at reviewing against criteria set in main.

Amount of interim assessment is trade off between complacency and helping people improve - would this need extra money into system?

With long cycle more of a worry about bringing people in just before and then declining. Could also put in Higher Education Statistics Agency data to trigger reassessment.

Implementation	What could go wrong?	Hoped for changes in UK research
Broaden units of assessment and use professional association to consult over make up of panels (make panel broader than at present but narrower than Research Councils). Subpanels – report to overall panel for different disciplines, provides flexibility. Panel determine criteria and publish (rules, pay offs, penalties, time frame). Universities submit for complete assessment, panels output detailed reports and criteria for upgrading. Interim self assessed criteria for upgrade if department wants to be reassessed.		Quality output appropriate to discipline from panels. More innovative research over longer period. Stability and improvement in research culture. Rewards different types of departments and universities. Provides funding councils more ammunition for greater funding from central government.

Discussion

There is no cross disciplinary comparability. This allows flexibility.

Hybrid System 17: Based on Expert Review

How do you find experts and recruit them?

Need much better system of using international assessors, need them involved from start.

Need to broaden international beyond North America.

How do you deal with interdisciplinary research? Make UoAs wider - but how do you get experts that can cover whole area? Same size panels but allow cross discipline workers to partially submit to another panel and get credit for this in primary panel.

Need to allow people to submit in other departments to their home departments.

End users beyond academia need to be included - eg civil servants for social policy research.

Want a longer time scale maybe 7-8 years, with voluntary assessment at half way. This should be an issue for the department rather than the university. This brings in an element of historical ratings as 5*'s would not want mid term assessment.

Need to measure other aspects of research culture eg nurturing of Post Docs vs employing contract research staff.

Identifying high flyers in low rated departments - liked the idea but thought it would be inefficient to give one good researcher lots of kit ie would trade off efficiency in some disciplines.

Game playing could be reduced by submission of all staff.

Wanted to use data from grant application process in some way.

Discussion

Use of lay panel members was discussed

Danger of bringing in new biases by using high court judges for example

Could also bring views of users to panels by other mechanisms than putting them on panels

Do end users have to be involved? Reworded as 'If there are end users they must be included'

Implementation	What could go wrong?	Hoped for changes in UK research
Assumed time span would be 8 years with half way voluntary.	Lack of consensus.	Improve UK research quality.
Consultation process by HEFCE to arrive at fewer UoAs.	Consultation overload.	Strengthen utility to end-users.
HEFCE and professional bodies consult and arrive at composition of panels and balance between peer/non-peer and end users).	Panels diverge too much on criteria.	Strengthen international peer assessment.
All panel chairs meet to arrive at common criteria.	New UoAs:	Promote interdisciplinary research.
Panels meet to propose specific criteria.	Lack of ownership/identification.	Sense of ownership by community.
Professional bodies/academic community is consulted on criteria.	Fails to solve interdisciplinary research problem.	

Discussion

A bit more top down - HEFCE makes initial decision and then consults on it. This could be a recognition and reality.

Hybrid System 18: Based on Expert Review

Key is to measure outputs of quality research.

Field dependent to accommodate research outputs and practice.

Keep number of UoAs but add a mechanism to increase value of interdisciplinary research.

Experts are mostly peers, some from outside academia and UK.

Use transparent and credible system to appoint panel members and chairs. Make use of professional bodies and academic societies (with regional agenda, ethnic and academic age, diversity all in set proportions).

Elements of self assessment should be more prospective - taking in aspiration and future strategy.

UoAs currently are graded as a whole, although submission is individually based. Want aggregated submission as UoAs then could abandon threshold for individuals. Probably longer time scale.

Need to look at research income but want to take into account productivity of grant income and efficiency and added value.

Also want to try and use algorithms to rate research culture and activities eg PhDs output etc.

Discussion

How do you measure added value? Liked the idea but think it hard to do. Have panel score against criteria.

Currently same output with more money gets better rating, this system reverses this which is an improvement.

Could there be more weighting to the aspirational and strategy? Could use a weighting system, maybe 60% for peer assessment and 15% for other two streams (strategy and research value added).

Implementation	What could go wrong?	Hoped for changes in UK research
Similar system to RAE2001 so there need to be relatively few changes.	How to balance consultation of community and panel autonomy.	More transparent method for selecting panel members. Better assessment of effectiveness of research income. Allow for more consistent approach and better fit to discipline.
Selection of panel members by HEFCE with professional bodies etc.	Difficulty of finding diverse panel members. Ambiguity in who is submitted.	
Need to identify a system of equivalence for submissions, use written explanation of selection criteria used by institution for inclusion in submission.	Complexity of assessment of appropriateness of outputs submitted.	
Devise algorithm using ranking that takes into account total volume of research funding in discipline and volume and types of outputs to produce a value for money measure.	Sub-disciplines may have different balances of outputs and value for money.	

Hybrid System 19: Based on Expert Review

Research needs a better classification system eg:

  Formal Sciences (Output Theorems)
  Explanatory Sciences (Output Laws)
  Design Sciences (Output Technological Rules)
  Human Sciences (Output is Artifacts and Knowledge)

All of these can be assessed using a combination of algorithms and expert review, combined in different and relevant ways.

Groups may report in more than one field.

Metrics can be discipline dependent (example for medicine/science).

Algorithm could include - citation, longevity, impact factor, research income, students and young staff.

Expert Review should moderate outcome of algorithm and include strategic judgments of research business plan, new initiatives and impact on users.

Discussion

Seeking to break down interdisciplinary boundaries completely.

Its the importance looking at the relevant outputs.

Implementation	What could go wrong?	Hoped for changes in UK research
Define fields - individual up rather than HEFCE down.	Mindset shift - could be confusing classification.	Assessment linked directly to purpose and output.
Many panels considering fewer and smaller submissions.	Lack of comparability across panels, but is this necessary?	Get rid of 'one size fits all' assessment.
Panels set up algorithms and measures.	Increased bureaucracy and admin at HEFCE.	Transdisciplinary - breaks down silos.
	How is the assessment linked to the money?	Tighter tying of funding to individuals leads to smoother graduation of cash?
		Increased dynamism returns to the system.

Hybrid System 20: Based on Algorithms and Metrics

System based on Metrics with moderation by Peer Review.

Variable applicability by discipline, which needs to be addressed.

Has advantage of being based upon outputs.

Expert panels would devise metrics for their fields and other mechanisms if metrics could not be devised.

A degree of self-assessment and historical performance but within very strict criteria to prevent flowering up.

Implementation

   Machine based data entry (to allow for automation).

   Criteria for evaluating discursive elements of self-assessment.

   Criteria governed by peer/expert group in discipline specific manner.

   Use a broader rating system that has finer grades so less step changes, this is allowed by having a metric based system.

   Should be caveat system for explaining individual's circumstances eg ill, child birth etc.

   Beefed up system of comparisons between areas of cognate disciplines.

   Criteria/principles:

      Impact factors.

      Recognition of excellence and ability to compensate for apparent under production.

      Metrics: funding income, citation, volume.

         Value for money.

Discussion

Such a system would force an extreme division between subjects in which metrics can be used and those which don't have suitable metrics.

In science judgment moderates data and in arts data moderates judgment.

Should be more comparison and dialogue between cognate subjects.

Implementation	What could go wrong?	Hoped for changes in UK research
Identify existing data sources and make use of these.	Poor data or data fabrication.	Maximizes use of existing data and other sources.
Appoint panel - including overseas and international experts, and these would need to be paid.	Inappropriate metrics.	Less distraction from core activity preparing assessments.
Identify appropriate metrics on discipline specific basis and decide importance of applicability.	Information overload.	Allows greater discrimination between different kinds of research.
Could all be submitted on generalized web based forms.	Inconsistency between UoAs.	More transparent and more explicit process.
Criteria for operation and evaluation, set detail of assessment system.	Instability in external environment influencing metrics and data.	
Cross-discussion between UoAs.		
Feedback and external audit.		

Hybrid System 21: Based on Expert Review

Panels with disciplinary focus, but cognate disciplines represented as well and end users as appropriate.

Transparent and clear criteria, process and feedback.

Assessment process that happens every 5-6 years.

With an output that has a developmental value/impact for submitting institutions.

A process that assesses current and potential research (not historical ratings).

Changes required:

    A process that uses documents that are already available rather than specially prepared for RAE eg publication lists, grants database, research strategy, HESA data.

    Use of institutional visits for broader sampling.

    Broader representation on panels.

Aiming for:

    Lighter touch.

    Feedback of benefit to institutions.

    Wider engagement of end users.

Discussion

Seconded idea of using documents that exists.

Very difficult to make objective assessment of potential. This system would also make university nurture better.

Principles are great - but there is contradiction of letting panels visit the institution and lighter touch. Want to capture what the institution is doing on an everyday basis, the reasoning behind visits.

Should the process provide directed feedback - what is that balance between assessment and supplying information for improvement?

Implementation	What could go wrong?	Hoped for changes in UK research
Select panel.	Damaged by self-serving behaviour.	End user focus.
Decision on flexibility and frequency.	Parity between areas more problematic.	Increased interdisciplinary work.
Consultation with subject community to decide balance between metrics and expert judgment.	Criteria may lead to loss of integrity - ie research chases rewards rather than excellence.	Less deadline driven.
Centers provide information and there is more flexibility in research outputs.	Vulnerable to information systems failure.	Encourage new areas - and allow more creativity and risk taking.
Panel members assess 'evidence' submitted eg CVs, research student completions etc.		
Test the systems to ensure data management systems work.		

Hybrid System 22: Based on Expert Review

Expert Review and Algorithms.

Weighting applied to each element should depend on discipline.

Maximize objectivity and use end user judgments where appropriate.

Disaggregate units and rate individuals.

Each department or institution could end up with a number of grades in different fields. —

Algorithm every 2 years plus expert panel at longer/self selected interval. Universities could enter groups for different modes of assessment.

Should be a way for institution to call in assessment for new units.

Outward looking criteria used to define research outputs.

External and International to be full members (possibly Chairs?) of panels.

Discussion

How do you define end user? A big problem with Theology. Subject areas should define their own end users and the impact they wish to have. But there is also the issue of finding users who could judge the usefulness of the research.

How do you rate utility for humanities? In humanities there is a particular problem with negative citation. 'Systematic reviews' in medicine are mainly synthetic but get rated highly by citation and are useful, but aren't really research.

Implementation	What could go wrong?	Hoped for changes in UK research
What data and how it will be collected?	Algorithm not applicable to given discipline.	More equitable distribution of money.
Some suggestions: Gather CVs in standardized form, bibliometric data, grant income, PhD completion.	Applicable but still unfair - might reflect fashion or marketability rather than excellence.	Increased credibility and improved external perception might increase external funding.
Need to devise weighting by discipline.	Honesty/citation clubs and joint authorships.	Increase capacity for rapidly recognizing positive change.
Identify experts and maybe use external referees and have panel adjudicate on reviews.	Composition of expert panel is biased or does not have expertise.	Encourage growth in institution and individuals ie get grades more swiftly if new.
Decide how to balance Expert Review and Algorithm.	Who picks the unit for people to submit into?	Reduce burden of assessment.
Choice of interval and when to call in panel for assessment (Could be done by shift in algorithm or by newly started units).		
Decide on form of output - more detail which allocates money by person so if they move they can take their money.		

Hybrid System 23: Based on Expert Review

They considered two alternatives and discussed merits of them: a light touch system and a heavier touch system.

Light touch system:

Include everybody and give them a score.

Use algorithm to allocate funding (could be a simple as on/off for Quality Related funding).

Couple this to larger funding of research councils (but need to add direct cost recovery to support Principle Investigators from research grants).

Need interdisciplinary comparison.

Heavier system:

Points are given for excellence of each paper, everybody is submitted, points are summed for each person and then across departments, then HEFCE funds the points.

Systems aim to address:

Bands that are too crude/coarse (so either refine grading or make on/off funding decision and easier to assess).

Declare everyone.

Avoid last minute transfers.

Bear in mind cost of panel evaluation.

Discussion

Put less money into RAE and so could use cruder system to allocate less money.

Pushing allocation of money over to Research Councils might actually produce more money spent on review/administration than for RAE, also would be useful to know how efficient the current RAE system is. Also need to decide what RAE is trying to do, is it trying to focus and improve research or is it about allocating a 'too small' pot of money to keep academics happy? Greatest importance is to recognize researchers of international standing - and work out how to develop these, should be far more emphasis on individuals rather than on institutional.

Implementation	What could go wrong?	Hoped for changes in UK research
		Cannot view research in isolation from teaching.
		We don't know what the impact of current RAE on research has been.
		Is bigger better? stronger? Is bigger feasible given dependence on teaching?
		Encourage collaboration? (maybe regional?)
		RAE should aim:
		To promote excellence in research.
		Legitimization/justification/trust of funding academic community with academics/politicians/public.
		Entire process is zero sum game, there is a limited pot of money.

Hybrid System 24: Based on Expert Review

Quantitative data review every two years.

Expert Review every five years with fewer larger panels.

Submit all staff.

Submit all outputs, self highlight four best.

Three grades for individuals: international, national, sub-national.

Clear criteria published early.

Independent chair of panel *from within cognate discipline but not from panel field.*

Funding would be on basis of number of individuals at below nation, national and international level.

Selecting panel members must have a transparent and accountable mechanism. There should not be self perpetuation.

Liked the idea of a panel chair from another area to avoid issues of self interest.

Panels should be about twenty people.

Panels should include a wider range of academics, don't need to be professorial rank, should come from wider range of institutions.

The rules of the game should be published as soon as possible. Once the consultation process is finished the criteria should be published and not changed.

International experts need better briefing and involved earlier in the process.

They liked the idea of directing funding directly to high-flyers in low rated departments and also the idea of a 'transfer fees' system.

Discussion

Would non-expert chair have credibility? Yes on a large panel where their role was to facilitate and convene.

Would a less elitist panel have a credibility in the field? Probably depends on the field.

Panel should not be selected on other issues than academic excellence, eg they shouldn't be selected for regional issues.

Implementation	What could go wrong?	Hoped for changes in UK research
Identify quantitative data on a rolling basis.	Implications for academic staff - would this force the introduction of 'teaching only' staff.	Assess quality of all academic staff.
Define outcome grades.	Unit of Assessment redesign consultative process.	Increase transparency.
Redefine UoAs (a complex and messy process).	Rolling matrix outputs rather than mixture of inputs and outputs (not quality).	Three quality of output based grades.
Create panels (long and difficult process).	Resource intensive.	Direct international input.
Publish criteria and modus operandi.	Fails to value interdisciplinary research.	Increase room for interdisciplinary research.
		Independent chair.

Discussion

Fewer grades? We already have three de facto grades in the RAE for funding so might as well recognize it.

Hybrid System 25: Based on Expert Review

Clear criteria that should be published early (and before panel appointed).

Selection of panel:

    Independent chair (not same discipline)

    Panel members characteristics:

        Broad expertise.

        Distinguished.

    Selection (up to 12, not bigger than at present, to facilitate decision making).

        Nominations by professional associations, research councils, universities.

        Selection by chair with advice from others including international experts and past chairs.

Changes to procedure:

    Citation measures calibration by seniority, by subject, by time-scale and by impact.

    International standing of panels need to be guaranteed - maybe each panel should have international member.

    Complete return of all staff.

    Returns of all publications and the four best should be highlighted.

    Clearly defined esteem indicators (with reference to RA6) provide list of acceptable features such as editorships, membership of Royal Society etc.

    Grants and research studentships should be reported by not decisive in judgments.

    5 Year period is about right as otherwise department condemned for too long, but recalibrate number of staff annually.

    Interdisciplinary research: have cross membership of panels, chair's responsibility to ensure it is taken into account.

    More feedback from panels.

    Comparison of distribution of scores across panels.

    Appeals process and ombudsperson.

Discussion

Some of these extra aspects can be already submitted as part of RA5 eg citation information, and additional evidence of publications past the 4 for each academics. Felt it was important to use citation information in a standard form if it was to be used.

A suggestion that unfair that people with 4 excellent articles get same score as someone with 8.

Implementation	What could go wrong?	Hoped for changes in UK research
Publish criteria now and identify units of assessment.	Refusal to serve on panels (might be linked to threat of litigation).	More innovative research if criteria are published early.
Appoint chairs.	Lack of funding to reward excellence.	Richer mentoring as need to bring all staff up to level as all staff entered into RAE.
Initiate search for panel members and appoint transparently.	Litigation.	Better interdisciplinary work.
Develop feedback mechanisms.	Relationship to research council.	Improve UK's international profile in research.
Formalize appeals procedure.	Introduction of two tier system.	Feedback allows progressive improvement.
	Boycott by institutions.	

Hybrid System 26: Based on Expert Review

Self-assessment is fed to panel and critiqued by panel, assessees then have opportunity to respond - this process would take a year.

All data from Expert Review, Self-assessment and *Algorithm fed through transparent process to produce much better feedback.*

Metrics are specific to units of assessment and set by panel, comparability assured by over arching panel.

Use current level of scrutiny.

Use a five or six year interval.

Additional funding for flagged areas of individuals within UoAs.

Funding consequences clear from the start, so cost benefit of grade enhancement by participation is clear.

Make a certain percentage of ranking eg top 10% should be 5*.

Discussion

Suggested that some fields are stronger in UK than others, so would be expected to have more 5*. Feeling that each discipline should have some flexibility in setting the percentage of departments considered 5*.

Implementation	What could go wrong?	Hoped for changes in UK research
Define rules and regulations governing process.	'Grade inflation' in self-assessment.	Improve status of lone workers with targeted funding.
Define panel members.	Might still fail to provide good feedback.	
Give panel members legislative boundaries to protect the panel - panel could be legal entity, so individual panel members cannot be sued.	Flawed panel composition.	
	Inappropriate balance of data and metrics.	
Define banding, percentages for each grade.	More steps could lead to more work.	
Define financial implications.		

Hybrid System 27: Based on Expert Review

Peer/Expert Review - with an increased role for Algorithms in a discipline specific way.

Criteria for metrics need to be clear and spelled out years in advance to facilitate targeting.

These criteria need to be tailored to each discipline, by a panel of experts (they would need to balance outputs, grant income, end user take up etc).

Metrics need reflect long term view to show the impact an individual or department has had on the subject area over the long term.

Use of metrics would reduce inconsistency in UoAs but allow differences between UoAs.

Increased use of metrics would reduce burden on panels and researchers.

Provide transparency of process and improved feedback system.

Provide oversight role with either overarching panel or 'shadow' panels.

Implement an improved appeals procedure - metrics would support this.

Fund highflying groups and individuals within low rated UoAs to reduce exit push.

Would prevent cheating by submitting in a UoA which is 'easier' than a 'home' unit, as you would have to submit within the UoA you had consulted in about the metrics.

Could include longer term assessment of impact of previous work (perhaps looking at impact of work submitted in RAE 1996 and 2001).

Discussion

Does issue of forcing people more rigidly into a panel have problems of provision, where a university doesn't have a department for a specific subject?

Implementation	What could go wrong?	Hoped for changes in UK research
Find a way to define metrics and their subject related balance.	Insufficient resolution in metrics for some subjects.	More emphasis on quality research and less emphasis on gamesmanship.
Define quality criteria as they apply to metrics.	Might increase gamesmanship based on playing metrics.	
Achieve panel representation.	Openness might lead to more challenges to system increasing burden.	
Deliver transparency, feedback and accountability.		
Ensure overarching comparability.		

Hybrid System 28: Based on Expert System

Series of points:

Started with Peer/Expert Review.

Some metrics may be universally appropriate (eg post grad submissions).

Other metrics to be included only if the specific panels want them eg income for some, citations for others.

Keep UoAs each with panel.

Broader panels to ensure comparability in broad subject areas (eg Engineering).

Size of panel should match size of discipline and UoA.

Refine grades and how levels of excellence are defined (for example local, national and international, what do these mean?).

4 rated units or below can opt out of the next round.

All or a high proportion of all staff should be submitted.

RA5/6 should be streamlined and shortened.

(both of the previous two should allow reduced burden on university).

There should be much more detailed and useful feedback.

For each person at least one piece of work (the one the individual considers best) should be read for everyone.

Discussion

Submission of all staff allows profile of staff and inclusion of young researchers could be counted as a good thing.

Does submission of all researchers disadvantage new disciplines, although flagging small groups could help address this point.

Could prioritization of different outputs be set on a discipline basis and how would you rate different levels of authorship?

Implementation	What could go wrong?	Hoped for changes in UK research
Identify criteria early on (in a discipline specific manner).	Panel chair could be inappropriate.	Would reduce grade inflation.
Establish rules for rates of staff return.	Get very narrow selection of panel and domination by old elite.	Would promote consistency across disciplines and between UoAs.
Identify panel members and chairs etc.	Government could renege on funding.	Would reduce game playing, by making it less attractive.
Clarify definitions of each level\grade and establish the scale (should there be a grade 6 - giving people a lower grade on the 'current' scale to make use of full the scale could be problematic?)	Efforts to explain system might fail to be made.	Would University of Ulster and Queens University Belfast merge?
	Would grades changes negate reform and cause grade inflation?	Demand for special system for Northern Ireland or a better way to protect research money?
	Universities would try to find ways to play games with new systems.	

Hybrid System 29: Based on Expert System

Key Elements:

Use Peer Review and Algorithms to provide both objectivity and flexibility.

Panel members:

Chair should have international credibility with chairing skills and no bias.

Others should be truly representative.

Measures:

Emphasis on outputs:

Use published works.

All submitted work should be peer reviewed.

Should have measures that look at research training.

Discussion

Independent chair would be very useful as a non-independent chair is unlikely to favor ideas that do not benefit their institutions.

The chair should simply be chairing and facilitating.

Need to involve international experts from the very beginning and engage them fully in the process.

Only allowing peer reviewed output to be submitted would help to cut down burden. But this might not be possible for all disciplines - eg curating or important policy reports.

International tends to be US, need to make this truly international and allow submission of best work that may not be in English, eg work in Russian on Russian politics.

If you are working in a field where you are the only person in the country there is no national level to compare yourself to, it is not a stepping stone on the way to international. Why is quality related to geographic distribution? The confusion of level of quality and geographical relevance. Academia is trans-national or anational so why use geographical excellence as a level.

Implementation	What could go wrong?	Hoped for changes in UK research
Define Grades (1-5*).	Disproportionate representation on panels.	Equitable funding across the UK system.
Define peers, size and composition of panels and chairs.	Grades unworkable in the face of the evidence.	Transparency of criteria.
Set up panels with chairs of panel.	Failure to declare bias or hidden criteria used by panels.	Improvement of panels leads to increased international credibility.
Panels define assessment criteria for their units.	Assessment criteria too general or too narrow.	More positive attitude towards assessment.
Define parameters of evidence (what do panels want to see?).	Too much for panels to read and consider.	Ownership of the system by researchers.

Discussion

Reinforced point about importance of ownership.